DOCUMENT RESUME

ED 167 625 TM 008 440

AUTHOR Griffiths, Alan K.; Cornish, Alan G.
TITLE An Analysis of Three Recent Methods for the Identification and Validation of Learning Hierarchies.
PUB DATE Mar 78
NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Comparative Statistics; Expectancy Tables; Foreign Countries; *Learning Characteristics; *Mathematical Models; *Organization; Psychometrics; Research Methodology; Senior High Schools; Sequential Learning; Statistical Analysis; *Statistical Studies; Validity
IDENTIFIERS Canada; Gagne (Robert M); Guttman Scales; *Learning Hierarchies

ABSTRACT

This paper reviews methods which have been used to identify and validate learning hierarchies. Specifically, three methods which reflect the notion of a relatively inviolate sequence are discussed and compared. Two of the psychometric methods compare pairs of skills, the ordering-theoretic method (Bart and Krus) and the test of inclusion (White and Clark). Dayton and Macready's method considers the hierarchy as a whole. Logical and empirical analysis with respect to the application of each of these methods to a common set of data (regarding a tenth grade chemistry concept) are reported. Results suggest that each method has some advantages. A modification of the White and Clark test, which enhances its generalizability, is also described. (Author/RD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED167625

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AN ANALYSIS OF THREE RECENT METHODS FOR THE IDENTIFICATION AND VALIDATION OF LEARNING HIERARCHIES

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Alan Griffiths

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Alan K. Griffiths and Alan G. Cornish
Memorial University of Newfoundland
St. John's, Newfoundland
Canada

Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada, March, 1978.

TM008 440

An Analysis of Three Recent Methods for the Identification and Validation of Learning Hierarchies

1. Introduction

The question of appropriate sequencing of instructional content is of obvious concern to educators. For almost two decades Robert Gagné's hierarchical model of learning has provided a major focus for attempts to answer this question.

Gagné (1973) describes two essential characteristics of learning hierarchies. Firstly, in the ideal case a learning hierarchy represents a set of intellectual skills which forms a pattern in which success at each skill is possible only for subjects who have mastered each subordinate skill. Secondly, the learning of subordinate skills contributes substantially to the learning of related superordinate skills. Reviews by Briggs (1968), Resnick and Wang, (1969), Walbesser and Eisenberg (1972), White (1973), Gagné (1973), and Cotton and others (1977) attest to the breadth of research which has been generated by this model. Despite this, relatively few hierarchies have gained general acceptance. In fact, this failure may be attributed to the difficulties encountered in developing effective measures of hierarchical dependance. The limitations of those measures used in the majority of reported studies have been succinintly expressed by White (1973). He suggests that typically errors of measurement have been ignored, that the statistical techniques used were inadequate, and that subjective judgements were usually made regarding the mastery status of the subjects. Further, he showed that application of three of the major indices suggested by Gagné and Paradise (1961), Walbesser and Eisenberg (1972) and Capie and Jones (1971) to the same data set led to very different results. Such criticisms are potentially devastating to the credence of the results emerging from this research. Clearly, real progress is possible only if better techniques become available.

The methods which have been used to validate learning hierarchies may be grouped into two classes; those which reflect the transfer properties of hierarchies and those which reflect the notion of a relatively inviolate sequence. We believe these to be of equal importance, and indeed the present paper is part of a larger study in which each is applied. However, in this paper we have chosen to focus only upon methods utilizing the second aspect. Examination of the literature since White's paper, indicates three more recent methods of this kind. Two of these methods, an 'ordering-theoretic' method (Bart and Krus, 1973; Airasian and Bart, 1975) and a 'test of inclusion' (White and Clark, 1973) focus upon comparison of pairs of skills. The third method (Dayton and Macready, 1976) considers the hierarchy as a whole.

In the present paper the characteristics of each of these three methods are discussed and compared. One of them, the White and Clark test, is extended to provide a more general model. The application of each of the three methods to the same data set is then described, and the paper concludes with suggestions for further practice.

2. The Ordering-Theoretic Method

When the validity of a hierarchy is determined by considering the relationship between pairs of elements, the essential nature of the problem may be illustrated by reference to a contingency table. In Figure 1 ✓ denotes possession and x non-possession of a skill, while A, B, C, D represent the observed frequencies in the appropriate cells. It is evident that high A and C values tend to be supportive of an hierarchical relationship while high D values tend to deny the relationship. Most of the attempts to identify hierarchical relationships have focussed on the loadings in such matrices.

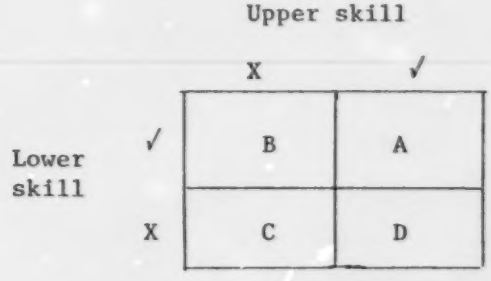

Figure 1

Most recently Bart and Krus (1973) and Airasian and Bart (1975) have taken this approach. Their method is described as an ordering-theoretic approach. It focusses upon the percentage of subjects whose responses are disconfirmatory to the existence of hierarchical connections between pairs of skills in a postulated hierarchy. In terms of Figure 1 the test is that the percentage of subjects in cell D should not exceed an arbitrary prespecified tolerance level. The test is applied to all possible combinations of pairs of skills in the hierarchy, including those not hypothesized to be hierarchical. A composite hierarchy is then identified. The method is deterministic, and does not take into account errors of measurement. No test is provided to determine the statistical confidence which can be attached to the existence of each accepted hierarchical relationship.

3. The White and Clark Test of Inclusion

A more sophisticated test for comparison of pairs of skills is given by White and Clark (1973). This test does allow for errors of measurement and also provides a test of statistical significance. Subjects are required to answer at least two questions per skill. In fact, White and Clark offer solutions only for the cases when both skills are tested by two questions or by three questions. [In section 6 of our paper we generalize the White and Clark procedure to any number of questions per skill, and for unequal numbers of questions per skill. We also make appropriate modifications to the significance test.] A skill-by-skill matrix of scores is then formed, the scores ranging from zero to two or three as appropriate. The matrix representing two questions per skill is shown in Figure 2.

Figure 2

Skill II (Upper) Questions Correct

Skill I (Lower) Questions Correct	0	1	2
2			
1			
0			X

The Critical Cell Is Marked X

The cell representing a score of zero on the lower skill and maximum possible on the upper skill is used to test the hierarchical relationship. This cell is assumed to contain those subjects most likely to possess the upper skill and lacking the lower one. The basis of the method is to test the null hypothesis that there will be no entries in this critical cell, other than those representing errors of measurement. The probability that the observed frequency does not violate the null hypothesis is calculated by using the marginal totals. For the case of two

questions per skill the probability that a member of the sample will be found in the critical cell is:

$$P_{02} = P_0 (1-\theta_b)^2\theta d^2 + P_I (1-\theta_a)^2\theta_d 2 + P_{II} (1-\theta_b)^2\theta_c{}^2 + P_B (1-\theta_a)^2\theta_c{}^2$$

where

P_0 = the proportion of the population with neither skill

P_B = the proportion of the population with both skills

P_I = the proportion of the population with skill I only

P_{II} = the proportion of the population with skill II only

θ_a = the probability of someone with skill I answering correctly any skill I question

θ_b = the probability of someone without skill I answering correctly any skill I question

θ_c, θ_d are the corresponding probabilities for skill II.

To make the estimate of P_{02} as large as possible and hence reduce the possibility of Type I error, θ_b is assumed to be zero and θ_d is assumed to equal one. That is, it is assumed that all subjects with one skill I question correct really possessed skill I and all those with one skill II question correct lacked the skill. For the case of three questions per skill appropriate modifications are made to the derivation above. A further modification suggested by Linke (1975) allows for the same procedures to be used with one and two percent exceptions in addition to those representing errors of measurement. This implies that resulting hierarchies represent substantial rather than absolute levels of hierarchical dependence.

4. The Dayton and Macready Model

The 'ordering-theoretic method' and the White and Clark test suffer from the disadvantage that they can only consider hierarchical connections between pairs of skills. From the results of these comparisons a composite hierarchy is produced. Such a procedure is less satisfactory than if the hierarchy could have been considered

as one entity. The Dayton and Macready model allows this, at least for small hierarchies. The basis of the Dayton and Macready model may be traced to Guttman (1944). Although he was not concerned with learning hierarchies, the simplest type of learning hierarchy (i.e., a non-branching, linear pattern) represents the form of a Guttman scale. For a perfect hierarchy (scale) no subject should exhibit a later skill if he fails to exhibit any earlier skill. Such responses constitute error. To maintain some standard for acceptance or rejection of the scale, and at the same time allow for some reasonable level of error, Guttman derived an index of 'reproducibility.' This was defined as the quotient of total errors (i.e., deviations from a perfect scale) over total responses, subtracted from one. Arbitrarily, a reproducibility of at least 0.90 was declared necessary if the hypothesized scale was to be considered valid.

Guttman's method has been criticized on several grounds (Lingoes, 1963). One obvious shortcoming is its limitation to linear scales or combination of linear scales. A second limitation is that the statistical tests are arbitrary. Proctor (1970) suggests them to be pre-statistical, and describes a method designed to elevate scaling to a better statistical foundation.

The basis of Proctor's model may be illustrated by the following example. Suppose for a three-item scale every subject in a population belongs to one of three Guttman true types which we denote by the vectors (0,0,0), (1,0,0), (1,1,0), (1,1,1), [where 0,1 represents non-possession and possession, respectively, of a skill and the hypothesized lowest skill is the first element and the highest skill is the last element in each vector.] In a test of the validity of the scale there will likely be subjects whose response patterns are not one of the four valid ones above. Under the null hypothesis such invalid responses are misclassifications. Proctor uses a misclassification parameter α to represent the probability of an invalid response to an item. If θ_1, θ_2, θ_3 are the proportions of the three types in the population then

the probability of a response of the type (0,1,0) for example is given by

$$Pr(0,1,0) = \alpha(1-\alpha)^2\theta_1 + \alpha^2(1-\alpha)\theta_2 + \alpha(1-\alpha)^2\theta_3 + \alpha^2(1-\alpha)\theta_4$$

Assuming a multinominal distribution of frequencies of the observed response patterns for a given sample, Proctor then obtains maximum likelihood estimates of the proportions of true types and the misclassification parameter by an iterative procedure, and determines the goodness of fit between data and model by a chi-square test. Proctor's suggestion that his model is of intrinsic interest is modest. Although it has not been directly applied in validation of learning hierarchies, it forms the basis for Dayton and Macready's (1976) intent to overcome the other main objection to Guttman scaling. That is, the Dayton and Macready method offers the possibility of extension to hierarchies of any configuration. These authors also capitalize on a suggestion by Proctor that it should be possible to allow separately for a '1-for-a-0' error and a '0-for-a-1' error. These they call 'guessing' and 'forgetting' parameters, α and β respectively.

Dayton and Macready's extension is discussed in some detail because it is one of the methods used in the present study. Consider, for example, the case of a five element hierarchy. Whatever the configuration of the hierarchy there are 32 possible distinct response patterns. If the hierarchy is linear only six of these are in agreement with the hierarchy. The remaining 26 would be considered misclassifications.

Now consider the following two branched hierarchies, each also containing five elements.

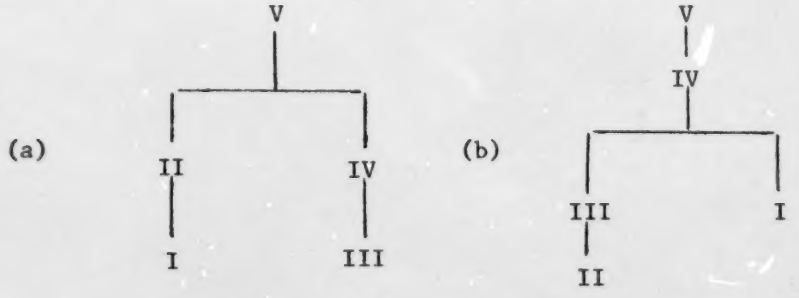

Figure 3

In (a) ten response patterns (00000), (10000), (11000), (00100), (10100), (11100), (00110), (10110), (11110), and (11111) are true to the hierarchy, with the remaining 22 patterns representing errors. In (b) there would be only eight true response patterns, namely (00000), (10000), (01000), (11000), (01100), (11100), (11110), (11111).

Using Dayton and Macready's (1976) notation, the probability of a subject producing a specific response pattern if the hierarchy is valid, is given by

$$\text{(i)} \qquad P(u) = \sum_{i=1}^{q} P(u/\nu_j)\ \theta_j$$

where

$$\text{(ii)} \qquad P(u/\nu_j) = \prod_{i=1}^{k} \alpha_i^{a_{ij}} (1-\alpha_i)^{b_{ij}} \beta_i^{c_{ij}} (1-\beta_i)^{d_{ij}}$$

Essentially what is implied by the use of these two equations is that the values of the misclassification parameters α and β are respectively raised to a power representing the number of 'guessing' and 'forgetting' corrections necessary to fit all 'true' pattern vectors to an observed data vector. Similarly $(1-\alpha)$ and $(1-\beta)$ are raised to the number of 'correct' responses in each case. The product of these over all possible response patterns is equation (ii). Multiplying this by the probability that the j^{th} true pattern vector occurs (θ_j), and summing this for all true pattern vectors yields equation (i), which represents Dayton and Macready's general probabilistic model.

Dayton and Macready discuss the use of their model with respect to three cases. In Case A, α and β are unrestricted. That is, there are separate guessing and forgetting parameters for each task, although the same parameters characterize a given task across different 'true' pattern vectors. Unfortunately Case A has been solved to date only for concept attainment models, defined by the authors as models in which each subject responds completely correctly or completely incorrectly to a given set

of tasks. While of potential use to those interested in criterion-referenced testing, at present Case A is therefore not of general use in validating learning hierarchies. In Case B all α's are the same for different tasks and all β's are the same for different tasks. In Case C, which is equivalent to Proctor's model, only one misclassification parameter is included. That is, α equals β. In all three cases maximum likelihood estimates of the various α's, β's, and θ's are obtained through a series of iterations. These values are used to compute the number of expected responses for each possible response pattern. The goodness of fit between data and hierarchy model is then calculated by both a Pearson chi-square test and a likelihood ratio expressed in the form of a chi-square. The latter appears to be more useful as it is less severely distorted by small frequencies, an important advantage when it is realized that for all except very small hierarchies a large proportion of the expected frequencies should be nearly zero if the hierarchy is valid.

It appears fair to conclude that the Proctor model, and hence Dayton and Macready's Case C, represents a major gain over the earlier scaling methods. Further, since the precision of Dayton and Macready's general probabilistic model is dependent upon the precision of the misclassification parameter(s), it is clear that their Case B represents an advance over Case C. In theory, Case A, in which the misclassification parameters are unrestricted, represents the most appropriate situation. Since Case A has so far been solved only for the relatively simple concept-attainment model, at the present time the most appropriate application of Dayton and Macready's model, and of scaling in general, to the determination of learning hierarchies is Case B. To date, there do not appear to be any reports of its use in applied situations.

5. Advantages and Disadvantages of the Three Methods

Of the three methods discussed the Dayton and Macready model is considered conceptually the most pleasing, because it considers the hierarchy as a whole--or at

least in larger pieces than pairs of elements. However, it has several important disadvantages. Firstly, the computer programme which is essential to its application can only accomodate small hierarchies at present. Secondly, incorrect response patterns are accommodated to the hierarchical model by means of a guessing parameter common to all elements in the hierarchy and a forgetting parameter similarly common to all elements in the hierarchy. The estimated values of these two parameters affect the predicted frequencies of *all* response patterns, thereby diminishing the potential precision of the model.

Of the two models involving comparisons of pairs of skills, that of White and Clark (1973) is easily the more sophisticated. The Linke (1976) modification of the test is potentially an improvement, although it needs more investigation. Dayton and Macready claim that their model subsumes that of White and Clark, and that the latter is equivalent to their Case A. Although true in theory this claim is misleading in practice, because Dayton and Macready have solved Case A only for the concept attainment model. This suggests, measurement errors excepted, that for each separate skill in the pair under test any subject should answer all questions correctly or incorrectly. This is not implied by White and Clark, and is less likely to be true as the number of questions increases. The problems caused by the use of common misclassification parameters in Case B and Case C of the Dayton and Macready model are avoided in the White and Clark test because the latter has, effectively, a guessing and a forgetting parameter for each skill. In this sense, rather than being subsumed by the Dayton and Macready model the White and Clark test has an advantage. Finally, Dayton and Macready correctly note that the White and Clark test is limited to equal numbers of questions per skill and to no more than three questions per skill. The last limitations we now remove in the following section.

6. Extension and Modification of the White and Clark Test

Consider again the problem of testing for pairwise comparisons of skills as

described on §3. Suppose now that subjects are required to answer m questions on the hypothesized lower skill and n questions on the upper skill, when $m,n \geq 2$. Using White and Clark's notation it is a simple matter to show that the probability Pij under the null hypothesis of a subject answering exactly i questions correctly on the lower skill and exactly j questions correctly on the upper skill is given by

$$(1)\quad Pij = \binom{m}{i}\binom{n}{j}\left\{\begin{array}{l} P_o\theta_b{}^i(1-\theta_b)^{m-i}\theta_d{}^j(1-\theta_d)^{n-j} + P_I\theta_a{}^i(1-\theta_a)^{m-i}\,\theta_d{}^j(1-\theta_d)^{n-j} \\ + P_{II}\theta_b{}^i(1-\theta_b)^{m-i}\,\theta_c{}^j\,(1-\theta_c)^{n-j} + P_B\theta_a{}^i\,(1-\theta_a)^{m-i}\,\theta_c{}^j(1-\theta_c)^{n-j} \end{array}\right\}$$

$i = 0,1, \ldots, m;\ j = 0,1, \ldots, n$. Summing (1) with respect to j, the expected number of subjects answering exactly i questions correctly on the lower skill is

$$(2)\quad N\sum_{j=o}^{n} Pij = N\binom{m}{i}\ \{Q\theta_a{}^i(1-\theta_a)^{m-i} + (1-Q)\theta_b{}^i(1-\theta_b)^{m-i}\} \qquad i = 0,1, \ldots, m,$$

where N = total number of subjects in the sample, and $Q = P_I + P_B$. Similarly, summing (2) with respect to i, the expected number of subjects answering exactly j questions correctly on the upper skill is

$$(3)\quad N\sum_{i=o}^{m} Pij = N\binom{n}{j}\{R\theta_c{}^j\,(1-\theta_c)^{n-j} + (1-R)\theta_d{}^j(1-\theta_d)^{n-j}\} \qquad j = 0, 1, \ldots, n,$$

where $R = P_{II} + P_B$. We denote by a_i, $i = 0,1, \ldots, m$, the observed number of subjects answering exactly i questions correctly on the lower skill and by b_j, $j = 0,1, \ldots n$, the number of subjects answering exactly j questions correctly on the upper skill.

Following a similar procedure to that of White and Clark (1973) we can obtain estimates $\hat{\theta}_a$, $\hat{\theta}_b$, $\hat{\theta}_c$, $\hat{\theta}_d$ of the parameters θ_a, θ_b, θ_c, θ_d. Equating, for example, the expected conditional probabilities of a subject answering m questions on the lower skill correctly given that he has m-1 or m correct obtained from (2), (3), to the estimate obtained from the observations, we have

$$\frac{a_m}{a_{m-1} + a_m} = \frac{Q\theta_a{}^m + (1-Q)\theta_b{}^m}{m[Q\theta_a{}^{m-1}(1-\theta_a) + (1-Q)\theta_b{}^{m-1}(1-\theta_b)] + [Q\theta_a{}^m + (1-Q)\theta_b{}^m]} \qquad m \geq 3.$$

Using the prior knowledge that $1-\theta_a$, θ_b, $1-\theta_c$, θ_d will be close to zero and ignoring terms in powers of $1-\theta_a$, θ_b, $1-\theta_c$, θ_d higher than the first we obtain, after rearrangement, the estimate

$$(4)\quad \hat{\theta}_a = \frac{ma_m}{ma_m + a_{m-1}}, \quad m \geq 3.$$

Similar arguments yield

$$(5)\quad \hat{\theta}_b = \frac{a_1}{a_1 + ma_0}, \quad m \geq 3.$$

$$(6)\quad \hat{\theta}_c = \frac{nb_n}{nb_n + b_{n-1}}, \quad n \geq 3.$$

$$(7)\quad \hat{\theta}_d = \frac{b_1}{b_1 + nb_0}, \quad n \geq 3.$$

We denote by $\bar{x}$, $\bar{y}$ the observed average numbers

$$\bar{x} = \frac{1}{N}\sum_{i=0}^{m} ia_i, \qquad \bar{y} = \frac{1}{N}\sum_{j=0}^{n} jb_j$$

of correct answers per subject on the lower and upper skills respectively. The corresponding expected values obtained from (2), (3) are, after simplification,

$$\frac{1}{N}\sum_{i=0}^{m} i\{N\sum_{j=0}^{n} Pij\} = m[Q\theta_a + (1-Q)\theta_b]$$

and $$\frac{1}{N}\sum_{j=0}^{n} j\{N\sum_{i=0}^{m} Pij\} = n[R\theta_c + (1-R)\theta_d]$$

Equating observed and expected average numbers and using (4) - (7) and the above approximations we obtain, after rearrangement, the estimates

$$(8)\quad \hat{Q} = \frac{\bar{x}/m - \hat{\theta}_b}{\hat{\theta}_a - \hat{\theta}_b}$$

and

$$(9)\quad \hat{R} = \frac{\bar{y}/n - \hat{\theta}_d}{\hat{\theta}_c - \hat{\theta}_d}$$

The estimates of P_I and P_O and P_B are related to $\hat{Q}$, $\hat{R}$ and $\hat{P}_{II}$ by the equations

$$\hat{P}_I = \hat{Q} - \hat{R} + \hat{P}_{II},$$

$$\hat{P}_O = 1 - \hat{Q} - \hat{P}_{II},$$

$$\hat{P}_B = \hat{R} - \hat{P}_{II}.$$

In the case of White and Clark's test, the test of significance of the null hypothesis H_0: $P_{II} = 0$ against the alternative H_a: $P_{II} > 0$ is performed by

first calculating an estimate Pon of $\hat{P}on$ by substituting the above estimates of the parameters (Under H_0) into (1). Denoting the observed frequency in cell (i, j) by fij, the significance of the observed data is given the by binomial expression

$$(10) \quad \text{sigdata} = 1 - \sum_{x=0}^{fon-1} \binom{N}{x} \hat{P}on^{x}(1 - \hat{P}on)^{N-x}$$

If sigdata is less than a prescribed significance level, H_0 is rejected. The White and Clark test depends, therefore, upon the observed and expected frequencies in the critical cell (o, n). When m and n are two or three this procedure is quite appropriate. However, when m or n exceeds three there is a case for increasing the number of critical cells. If S denotes the set of critical cells chosen as appropriate by the experimenter then the significance of the observed data is given by the multinomial expression

$$(11) \quad \text{sigdata} = 1 - \Sigma \frac{N!}{No!} (1-\hat{p})^{No} \prod_{(i,j)\in S} \frac{\hat{P}ij^{xij}}{xij!}$$

Where Σ is taken over all combinations of non-negative integers xij such that

$$o \leq \sum_{(i,j)\in S} xij < \sum_{(i,j)\in S} fij,$$

and $$No = N - \sum_{(i,j)\in S} xij,$$

and $$\hat{p} = \sum_{(i,j)\in S} Pij.$$

It is readily seen that (11) reduces to (10) in the case of a single critical cell (o,n).

As an illustration, in the case of m = n = 4 observations falling in the cells (0,3), (0,4), (1,3), (1,4) might be considered as evidence against H_0. These critical cells are shaded in figure 4.

Skill II (upper) questions correct

Skill II (lower) questions correct	0	1	2	3	4
4					
3					
2					
1				▨	▨
0				▨	▨

Figure 4

In the application in §8 of our extended and modified White and Clark test we use two, four and eight test items per skill. The critical cells we have chosen as appropriate to our experiment are as follows:

m = 2, n = 2: (0,2)
m = 2, n = 4: (0,3), (0,4)
m = 2, n = 8: (0,6), (0,7), (0,8)
m = 4, n = 2: (0,2), (1,2)
m = 4, n = 4: (0,3), (0,4), (1,3), (1,4)
m = 4, n = 8: (0,6), (0,7), (0,8), (1,6), (1,7), (1,8)
m = 8, n = 2: (0,2), (1,2), 2,2)
m = 8, n = 4: (0,3), (0,4), (1,3), (1,4), (2,3), (2,4)
m = 8, n = 8: (0,6), (0,7), (0,8), (1,6), (1,7), (1,8), (2,6), (2,7), (2,8).

In the case where one of m or n equals two the reader will not find it difficult to verify that the above theory holds almost unchanged. If m = 2 and n > 2, equations (4) - (9) are unchanged except that we must replace (5) with $\hat{\theta}_b = 0$. If m > 2 and n = 2, equations (4) - (9) are again unchanged except that we must replace (6) with $\hat{\theta}_c = 1$. (The rationale that White and Clark use to put $\hat{\theta}_b = 0$, $\hat{\theta}_c = 1$ in their case of m = n = 2 may be suitably extended in our case.)

There are two advantages to our extension of the White and Clark test. The first is that situations often arise in practice when three test items are not considered sufficient to assess mastery of a skill. Secondly, it is not always possible to ensure equality of m and n. (Sometimes, for instance, test items must be later rejected as unsuitable.)

With the extension of the procedure, inevitably, a modification is required of the binomial test to a multinomial test as we have described. Although the increased calculation is more tedious by hand a computer program has been developed (R.A. Cornish, 1977) to handle all values of m, n and all choices of critical cells.

7. Method

In the present paper we are not concerned with the particular content of

the hierarchy involved, but with the application of three methods to the same content. Therefore, the skills involved will not be described, but it may be noted that they represent development of an introductory chemistry concept, the mole concept.

An hypothesized hierarchy was derived by essentially following White's (1974a) model. Briefly, the pinnacle of the hierarchy was defined in behavioral terms. The hierarchy was developed downwards by asking the question of successively hypothesized skills "What must the subject be able to do in order to learn this new skill?" The hierarchy was modified after a check by teachers and subject matter experts, and was again modified following a pilot study. Contrary to White's suggestions, instruction was carried out directly by teachers rather than through a programmed format. Seven teachers and 209 Grade Ten chemistry students from Calgary public schools were involved. All testing and scoring was carried out by one of us, and one skill was tested by two items, two by four items, and four by eight items. The data in the present study was derived from testing at the end of instruction. While the question of what constitutes the appropriate time to test is a subject of some controversy, it is not relevant to the intent of the present paper. The data was analyzed by each of the three methods described. A computer program was written (Cornish, R., 1977) to facilitate the analysis of data by the modified White and Clarke procedure. Dayton and Macready's (1976) own program was used for that part of the analysis relating to their model. For the application of the Bart and Krus method, the data was cast into the form of a series of crosstabulation matrices representing all possible paired comparisons of skills, using the SPSS (1975) statistical package.

8. Application of the White and Clark Test

The results of applying our modification of the White and Clark test to our data are given in Table 1. The test is applied to all pairs of skills where the data suggested the possibility of an hierarchical connection, whether or not it was

hypothesized originally. Hence in the table the designation 'upper' and 'lower' does not reflect any theoretical position. Each cell in the table contains the number of exceptions to the particular hierarchical connection (that is, the total number of entries in the critical cells), and in parentheses the level at which the connection is considered valid. With regard to this, three levels only are considered. These allow for entries which can be ascribed to errors of measurement only (00), errors of measurement plus one percent exceptions (01), and errors of measurement plus two percent exceptions (02). In all cases a five percent level of significance is used. Where a connection is not validated at any of the three levels, only the number of exceptions is recorded. The skills are labelled A to H, skill A being the pinnacle of the hypothesized hierarchy.

Table 1

Application of the White and Clark Test at Three Levels of Stringency

'Lower' Skill	'Upper' Skill A	B	C	D	E	F	G	H
A	---	16	4 (00)	12	45	73	27	73
B	21	---	7 (02)	3 (00)	57	82	28	84
C	27	26	---	19	79	106	46	107
D	14	4 (01)	1 (01)	---	43	61	18	60
E	0 (00)	4 (01)	1 (01)	3 (01)	---	16	7 (02)	18
F	0 (00)	0 (00)	0 (00)	0 (00)	0 (00)	---	0 (00)	1 (01)
G	8 (02)	8 (02)	2 (01)	3 (00)	23	35	---	33
H	3 (01)	3 (01)	0 (00)	2 (00)	4 (01)	6 (02)	1 (00)	---

In terms of the hierarchical relationship which emerges, the results in

Table 1 may be more informatively expressed as in Table 2. The latter indicates those skills judged to be subordinate to each of the other seven skills, at each of the three levels of stringency applied.

Table 2

Summary of Hierarchical Corrections Identified
After Application of the White and Clark Test

Level	A	B	C	D	E	F	G	H
00	E F	F	A F H	B F G H	F		F H	
01	E F H	D E F H	A D E F G H	B E F G H	F H		F H	F
02	E F G H	D E F G H	A B D E F G H	B E F G H	F H	H	E F H	F

In determining the hierarchies that emerge from our data only those hierarchies leading to skill A as superordinate were considered, since this was the desired capability in this study. Depending on the stringency applied to the White and Clark test, three alternative hierarchies emerge. In order of decreasing stringency these are represented in Figure 5.

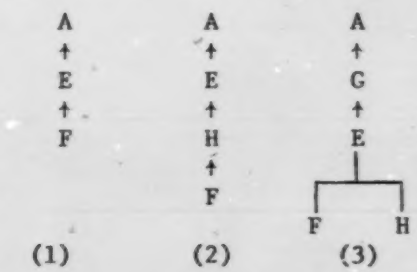

Figure 5: Hierarchies one to three

Consideration of these hierarchies raises several questions. Firstly, given three hierarchies to choose from, which is most appropriate? It would appear that the more stringent the test the more certain one can be of the validity of the hierarchy. However, this is not necessarily so. Some relationships appear to be only marginally established at one level because they disappear at a less stringent level, the relationship becoming bi-directional. For example, skill H is not found in Hierarchy One, appears to be superordinate to skill F in Hierarchy Two, and in the third hierarchy is equivalent to skill F. A similar relationship is found for B and D, although these do not feature in these three hierarchies. In general it may be suggested that larger hierarchies are more useful than smaller ones because they contain more information. However, while this might suggest testing at much less stringent levels, such a procedure would not be particularly useful. Taken to the ultimate extreme, the relationship between the skills in each pair would be validated in both directions. Hence there would be no hierarchy. At the opposite extreme only exceptions attributable to error of measurement would be allowed. This represents the absolute notion of an hierarchy, but may result in very small hierarchies of little practical use. As suggested by Gagné (1970) substantial rather than absolute levels of hierarchical dependence seem desirable, but what constitutes an optimum level is less clear. It is suggested, in the absence of better criteria, that it is appropriate to determine the hierarchies which emerge under progressively less restrictive criteria, indicating the hierarchies and the allowed percentage of exceptions in each case. When the number of uni-directional hierarchical connections begins to decrease analysis should cease. In the present study this occurs when analysis is continued beyond the 02 level.

In the hierarchies considered above some of the connections may be considered tenuous. In Hierarchy Two, skill H is superordinate to skill F, whereas in Hierarchy Three they are equivalent. The difference between them is not substantial.

Hence Hierarchy Four, represented below, is considered a reasonable alternative to Hierarchy Two.

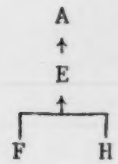

Figure 6: Hierarchy 4

A second potential anomaly relates to the position of skill G, which is not subordinate to skill A until the 02 level. In addition it is not superordinate to skill E until the 02 level, and it is difficult to justify its superordinate position relative to skills E and F in terms of the logic of the subject matter. White (1974a) suggests that connections which are illogical in the context of the subject matter should be ignored. We find it difficult to accept this position. Nevertheless, given the marginal nature of the connections discussed, we are prepared to entertain Hierarchies Five to Eight as possible alternatives to Hierarchy Three. Fortunately, the Dayton and Macready model allows us to test if these alternatives are significantly different. We return to this in §11.

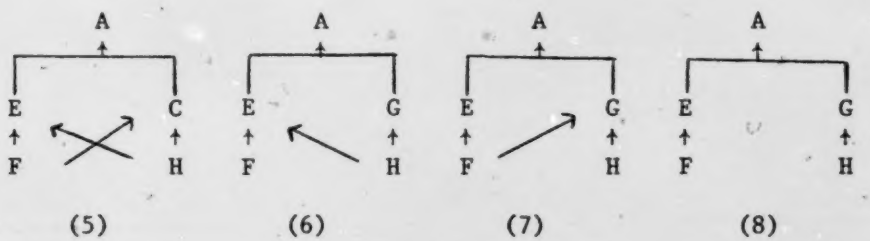

Figure 7: Hierarchies 5 to 8

9. Application of the Ordering-Theoretic Method

The greater precision of the White and Clark test compared to the use of the ordering-theoretic method has already been discussed. However, the recent popularity of the latter suggests that a comparison of the results obtained from the

application of each to the same data is appropriate. Moreover, the ordering-theoretic method is easier to use, and if it produces similar results to the White and Clark test it should still be useful. The classification of subjects as masters or non-masters of a skill is a problem in this study, as in others. For the application of both the ordering-theoretic method and the Dayton and Macready method, such classification is necessary. We have therefore performed each analysis twice, first using the complete sample and then on the same sample restricted by eliminating subjects with central scores. Hence, in the first instance, scores of at least one, three and six points were taken to represent possession and lesser scores to represent absence of a skill tested by two, four and eight items respectively. In the second instance scores of at least two, three and six were taken to represent possession of the same skills, while scores of less than one, two and three respectively represented lack of these skills; subjects neither classified as possessing or lacking these skills were dropped from the analysis. The intent of this second classification was to increase the certainty of classification, and hence confidence in the results of the application of the Dayton and Macready method to the particular data. A second possibility, which cannot be summarily discussed, is that different hierarchies may exist if one considers partial possession of skills as opposed to possession and non-possession.

Hierarchical corrections are tested for both the unrestricted and the restricted sample, allowing for one, two, and five per-cent exceptions respectively. The results are indicated in Tables 3 and 5. As for the White and Clark test all possible connections are considered. Tables 4 and 6 summarize the connections identified from consideration of the results shown in Table 3 and 5. (See pages 21-22.)

10. <u>The Level of Agreement Between the Results of Applying the White and Clark Test and the Ordering-Theoretic Method</u>

As a measure of the agreement between the results of applying the White and

Table 3

Ordering-Theoretic Method
Percentage of Exceptions to Hierarchical Connections, Complete Sample

'Lower' Skill	'Upper' Skill A	B	C	D	E	F	G	H
A		11	4.5	8.7	28	42	18	42
B	14		2.8	7.4	35	46	21	45
C	18	17		7.0	42	56	26	55
D	11	6.9	2.0		18	46	17	49
E	1.3	4.0	1.2	3.7		18	6.5	19
F	0.0	0.4	0.0	0.4	3.0		1.6	7.5
G	8.0	7.1	2.0	4.0	24	33		32
H	1.2	1.2	0.4	0.8	1.6	2.4	0.4	

Table 4

Summary of Hierarchical Connections Derived by Applying the Ordering-Theoretic Method (Complete Sample)

Level	A	B	C	D	E	F	G	H
1%	F	F	F	F				
			H	H				
2%	E	F	D	F	F		F	
	F	H	E	H			H	
	H		F					
			G					
			H					
5%	E	D	A	E	F	H	F	
	F	E	B	F	H		H	
	H	F	D	G				
		H	E	H				
			F					
			G					
			H					

Table 5

Ordering-Theoretic Method

Percentage of Exceptions to Hierarchical Connections, Restricted Sample

'Lower' Skill	'Upper' Skill A	B	C	D	E	F	G	H
A		8.9	2.2	7.2	25	39	17	40
B	12		3.7	1.8	32	46	18	44
C	15	14		12	44	56	29	55
D	8.4	2.4	0.6		27	36	12	36
E	0.0	2.2	0.6	1.9		8.2	4.4	9.6
F	0.0	0.0	0.0	0.0	0.0		0.0	0.5
G	5.0	5.0	1.3	2.0	15	20		21
H	1.6	1.6	0.0	1.2	2.1	3.1	0.6	

Table 6

Summary of Hierarchical Connections Derived By Applying the Ordering-Theoretic Method (Restricted Sample)

Level	A	B	C	D	E	F	G	H
1%	E	F	D	F	F		F	F
	F		E				H	
			F					
			H					
2%	E	F	D	B	F		G	F
	F	H	E	E			H	
	H		F	F				
			G	G				
			H	H				
5%	E	D	A	B	F	H	E	F
	F	E	B	E	H		F	
	G	F	D	F			H	
	H	G	E	G				
		H	F	H				
			G					
			H					

Clark test and the ordering-theoretic method to our data, Cohen's K (Light, 1973) was applied. The results are given in Table 7.

Table 7

Cohen's K as a Measure of Agreement between the Ordering-Theoretic Method and the White and Clark Test

		White and Clark		
		00	01	02
Ordering-Theoretic - (unrestricted)	1%	.61	.31	.21
	2%	.66	.69	.53
	5%	.45	.85	.86
Ordering-Theoretic - (restricted)	1%	.68	.56	.42
	2%	.64	.85	.68
	5%	.42	.82	1.00

Examination of Table 7 suggests that, irrespective of whether it is applied to the restricted sample or the whole sample, use of the ordering-theoretic method at the one percent level is most similar to the White and Clark test at the 00 level. At the 2% level, it is most similar to the White and Clark test at the 01 level. Finally, at the 5% level, it is most similar to the White and Clark test at the 02 level. A further finding is that the results from the White and Clark test are more consistent with those from the ordering-theoretic method for the restricted sample than with the unrestricted sample at each level of stringency. This better fit could be attributed to greater precision in classifying subjects, or alternatively to greater congruence between the implicit mastery criterion in the White and Clark test and the mastery criterion for the restricted sample when ordering-theoretic method is applied relative to that used when the sample is unrestricted. The complete agreement between the results of applying the test on the restricted sample at the five percent level and the White and Clark test at the 02 level encourages the interpretation that results from application of these two tests may be used interchangeably. The measure of agree-

ment observed in Table 7 for the comparisons at the other levels is similarly encouraging. However, the absolute number of differences observed when these tests are applied to the same data in the present study suggests the need for caution. Considering the use of the ordering-theoretic method on the restricted sample and the White and Clark test, six, four, and zero different connections were observed at the three successive stringency levels. Using the unrestricted sample for the ordering-theoretic method in similar fashion six, seven, and four differences were observed. In the present study little or no difference would have been found in the hierarchies suggested at each level if the ordering-theoretic method, especially for the restricted sample, had been used in place of the White and Clark test. However, this may have been fortuitous because many of the deviant results related to skill B which did not feature in the eventually preferred hierarchies. Nevertheless the results, particularly at the least stringent level, are at least quite suggestive that the ordering-theoretic method may be substituted for the White and Clark test. Whichever of these two tests is used the fact remains that the resultant hierarchy is a composite and is not tested directly. The Dayton and Macready method tests validity of an hierarchy as a complete entity. In the next section it is applied to determine the goodness of fit of Hierarchies One to Eight to our data.

11. Application of the Dayton and Macready Method

The Dayton and Macready method was described in detail in §4. The goodness of fit between data and an hypothesized hierarchy is determined by both a chi-square and a likelihood ratio expressed as a chi-square. In the present study the latter was used. Both statistics suffer from the fact that a substantial proportion of the expected frequencies will approximate zero. Hence the chi-square value may be substantially affected by as little as one response which is "false negative", that is, when the subject appears to lack a skill relatively low in the hierarchy while

exhibiting higher skills. Similarly, if the subject uinely lacks the lower skill but has an alternate but unidentified route to higher skills, the chi-square will be substantially affected. The likelihood function is less seriously affected. A "false positive" result, in which a lower skill appears to have been gained when it has not, is much less likely to distort the statistical tests. Hence use of the Dayton and Macready method is more likely to produce a type I error than a type II error. It is more likely that valid hierarchies will be missed than that invalid hierarchies will be validated. Consequently, correct classification of subjects as masters or non-masters of each component skill is particularly important. For this reason the Dayton and Macready method was applied using the same restricted sample as for the application of the ordering-theoretic method, as well as for the complete sample.

In computing the value of the likelihood ratio the Dayton and Macready method provides estimates of the guessing and forgetting parameters needed to provide a fit between data and hierarchy. As these values increase, confidence in the particular hierarchy decreases. They may be used to aid differentiation between alternative hierarchies, although the primary statistic used in this context is the difference between likelihood ratios. Table 8 and Table 9 indicate the likelihood ratios and misclassification parameters for each of the above hierarchies, in the two situations described. (See page 26.)

In Tables 8 and 9 the null hypothesis under test is that there is no difference between the observed frequencies and those expected if the hierarchy in question is considered true. For the unrestricted sample none of the three hierarchies (one, two and three) suggested after applying the White and Clark test was consistent with the data when the Dayton and Macready test was applied at the five percent level. However, Hierarchies Four and Seven, suggested as alternates to Hierarchies Two and Three respectively, were consistent with the data when the same test was applied.

Table 8

Likelihood and Misclassification Parameter Estimates for Hierarchies One to Eight (Unrestricted Data)

Hierarchy	Guessing Parameter Estimate	Forgetting Parameter Estimate	Maximum Likelihood Estimate	Degrees of Freedom	Significance of Data
1	.50	1.00	336	2	0.000
2	.00	.06	47.6	9	0.000
3	.22	.00	51.9	23	0.001
4	.28	.04	15.2	8	0.055
5	.20	.00	38.1	22	0.018
6	.20	.00	38.3	21	0.012
7	.18	.00	31.6	21	0.064
8	.19	.00	31.8	20	0.046

Table 9

Likelihood and Misclassification Parameter Estimates for Hierarchies One to Eight (Restricted Data)

Hierarchy	Guessing Parameter Estimate	Forgetting Parameter Estimate	Maximum Likelihood Estimate	Degrees of Freedom	Significance of Data
1	.00	.00	0.00	2	1.000
2	.00	.01	10.6	9	0.304
3	.28	.00	18.7	23	0.719
4	.00	.01	7.9	8	0.443
5	.23	.00	16.05	22	0.813
6	.23	.00	16.10	21	0.764
7	.20	.00	13.63	21	0.885
8	.20	.00	13.63	20	0.849

Further, another alternate to Hierarchy Three (Hierarchy Eight) was only marginally inconsistent at the five percent level. It appears, therefore, that the Dayton and Macready method is empirically superior to the other two methods considered in that it does allow an effective comparison of alternate hierarchies. It is of note here that in our study the preferred hierarchy following the Dayton and Macready analysis is not the same as that which emerges when the skills are considered in pairs. Nevertheless, tests such as the White and Clark test and the ordering-theoretic method are necessary to produce an initial sifting of the data so that hierarchies which are consistent for each individual connection may be generated as potential hierarchies for the Dayton and Macready method.

The question of whether Hierarchy Seven or Hierarchy Four is more appropriate is perhaps best answered on utilitarian grounds. In general we suggest that given the choice of two hierarchies which are both consistent with the data, the hierarchy containing more skills should be chosen because it contains more information. In this case, therefore, Hierarchy Seven appears to be the more appropriate. The superiority of Hierarchy Seven over the other five-skill alternatives may be further tested by examining the difference between its likelihood ratio from the Dayton and Macready test, relative to each of the others. The results of applying this test are given in Table 25. Δ represents the difference between the likelihood functions. The significance of the difference is determined by reference to a chi-square table.

Table 10

Likelihood Ratio Differences for Hierarchy Seven Against Four Alternatives (Unrestricted Data)

Alternative Hierarchy	Δ	df	Significance of Data
3	20.3	2	.000
5	6.5	1	.011
6	6.7	0	
8	0.2	1	.655

The results shown in Table 10 confirm that Hierarchy Seven is more appropriate than Hierarchies Three, Five and Six. Very little difference is observed, however, between it and Hierarchy Eight.

Table 9 shows that when the sample was restricted by eliminating those subjects for whom there was reasonable doubt about mastery status for any skill included in the hierarchy under test, all eight hierarchies (including the three derived directly from application of the White and Clark test) were consistent with the data. The most consistent hierarchy was Hierarchy One, which formed a perfect fit. Hierarchy Four was again superior to Hierarchy Two, although the difference in this instance was significant only at the ten percent level. In all three of these hierarchies the values of the misclassification parameters were negligible. In the remaining five hierarchies, all containing five elements, the guessing parameter was large (.20 to .28) although the forgetting parameter was again minimal. Again Hierarchy Seven provided the best fit. However, the difference between the value of this statistic for Hierarchy Seven relative to each of the others was never significant at the five percent level with the possible exception of that for Hierarchy Six. Hierarchy Three was weaker than Hierarchy Seven at about the eight percent level. Moreover the guessing factor was worst for Hierarchy Three. Hierarchy Eight was again very similar to Hierarchy Seven in its goodness of fit to the data.

In summary we suggest that the most appropriate hierarchy of the eight tested is Hierarchy Seven, because it offers a good fit to the data and is the best of the larger hierarchies to do so. Nevertheless, the large value of the guessing parameter is a concern. A smaller hierarchy, Hierarchy Four, also offers a good fit, and has negligible values for its misclassification parameters. However, the additional information given in Hierarchy Seven suggests the latter to be more appropriate. Whether it is in fact more appropriate than Hierarchy Eight is open to some question. It may be best to compromise here and to include the additional connection in Hierarchy

Seven over Hierarchy Eight as tentative. This may be represented in Hierarchy Nine:

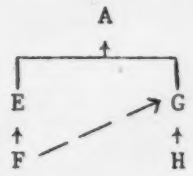

Figure 8: Hierarchy 9

A further amendment may yet be made. From Table 1 it is evident that skill C appears to be superordinate to skill A even at the 00 level when the White and Clark test is applied. It is not possible to justify this result in terms of the logic of the connection. Further, skill D is seen to be equivalent to skill A, a result which again is surprising in view of the particular subject matter. However, further examination of the subjects' responses suggest the possibility of a disjunctive relationship, as in Hierarchy Ten.

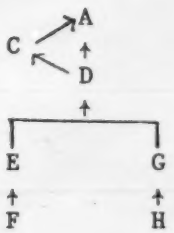

Figure 9: Hierarchy 10

Hierarchy Ten implies that some subjects may proceed to skill A via skill D directly, while others may proceed to skill A via skill D followed by skill C. These alternate routes have the effect of making skill C appear more difficult relative to skill A, and to a lesser extent it is possible that the same may be said for the relationship between skills A and D. The White and Clark test does not allow the identification of such a disjunctive relationship, while the other two methods do. Applying the Dayton and Macready test to the unrestricted data a likelihood function of 55.67 with

52 degrees of freedom is obtained. The guessing parameter is 0.18 and the forgetting parameter 0.02. Hierarchy Ten is therefore judged to be consistent with the data.

12. Conclusions and Recommendations

As we have indicated in §5 the Dayton and Macready method has two major advantages over the other two methods considered in this paper, and indeed over any other methods with which we are familiar. First, that it considers the hierarchy under test as one entity, even when that hierarchy is branched. Second, that it allows a direct test between hierarchies which contain the same skills in different arrangements. A further advantage, relative to the White and Clark test is that it can accommodate disjunctive connections within the hierarchy. We remind the reader, however, that to reach its full potential the Dayton and Macready model requires further elaboration. In particular the provision of separate 'guessing' and 'forgetting' parameters for each skill, would enhance the precision of the model.

Despite the advantages of the Dayton and Macready method an initial sifting of the data is often likely to be necessary before the method can be applied. What is implied here is that hierarchies hypothesized prior to collection of data are seldom completely substantiated. Methods such as the White and Clark test or the ordering-theoretic method are more flexible in identifying alternate hierarchies. These can then be tested as the hypothesized hierarchies in the application of Dayton and Macready's method.

We have indicated that the White and Clark test is likely to yield more reliable results than application of the ordering-theoretic method. Nevertheless, our own empirical evidence suggests that the difference may not be substantial. We consider this result to be quite suggestive, but recommend further research.

We have suggested a modification of the White and Clark test to allow for more questions per skill and for different numbers of questions per skill. This allows greater certainty in the identification of hierarchical connections, and also allows rejection of items found to devient.

Finally, we agree with the notion of substantial rather than absolute hierarchical dependance. In the absence of better criteria to define 'substantial' we recommend, with respect to the application of the ordering-theoretic method and the White and Clark test, that analysis continue at successively less stringent levels until the number of hierarchical connections decreases. However, the particular needs of the individual researcher may dictate otherwise.

References

Airasian, P.W.& Bart, W.M. Validating a priori instructional hierarchies. *Journal of Educational Measurement*, 1975, 12(3), 163-173.

Bart, W.M. & Krus, D.J. An ordering-theoretic method to determine hierarchies among items. *Educational and Psychological Measurement*, 1973, 33, 291-300.

Briggs, L.J. *Sequencing of instruction in relation to hierarchies of competence.* Pittsburgh, Pennsylvania: American Institute For Research, 1968.

Capie, W. & Jones, H.W. An assessment of hierarchy validation techniques, *Journal of Research in Science Teaching*, 1971, 8, 137-147.

Cornish, R. A computer program for the generalized White and Clark test (personal communication).

Cotton, J.W., Gallagher, J.P., & Marshall, S.P. The Identification and Decomposition of Hierarchical Tasks. *American Educational Research Journal*, 1977, 14(3), 189-212.

Dayton, C.M. & Macready, G.B. A probabilistic model for validations of behavioral hierarchies. *Psychometrika*, 1976, 41(2), 189-204.

Dayton, C.M. & Macready, G.B. Computer programs for probabilistic models, University of Maryland, 1976.

Gagné, R.M. Some new views of learning and instruction. *Phi Delta Kappan*, 1970, 51(9), 468-472.

Gagné, R.M. Learning and instructional sequence. In F. Kerlinger (Ed.), *Review of Research In Education (1)*. Itasca, Illinois: F.E. Peacock, 1973.

Gagné, R.M. & Paradise. M.E. Abilities and learning sets in knowledge acquisition. *Psychological Monographs*, 1961, 75 (whole No. 518).

Guttman, L.A. Basis for scaling qualitative data. *American Sociological Review*, 1944, 9, 139-150.

Light, R.L. Issues in the analysis of qualitative data. In R.M.W. Travers (Ed.) *Second Handbook of Research on Teaching*, Chicago: Rand McNally, 1973.

Lingoes, J.C. Multiple scalogram analysis: A set-theoretic model for analyzing dichotomous items. *Educational and Psychological Measurement*, 1963, 23(3), 501-524.

Linke, R.D. Replicative studies in hierarchial learning of graphical interpretation skills. *British Journal of Educational Psychology*, 1975, 45, 39-46.

Nie, N.H. & others. Statistical package for the social sciences (2nd. ed.) New York: McGraw-Hill, 1975.

Proctor, C.H. A probabilistic formulation and statistical analysis for Guttman scaling. *Psychometrika*, 1970, 35, 73-78.

Resnick, L.B. & Wang, M.C. *Approaches to the validation of learning hierarchies.* Educational Testing Service, 1969.

Walbesser, H.H. *An evaluation model and its application.* Washington, D.C.: American Association for the Advancement of Science, 1968.

Walbesser, H.H. & Eisenberg, T.A. *A review of research on behavioral objectives and learning hierarchies.* (ERIC Document 059-900), 1972.

White, R.T. Research into learning hierarchies. *Review of Educational Research*, 1973, 43(3), 361-375.

White, R.T. A model for validation of learning hierarchies, *Journal of Research in Science Teaching*, 1974, 11(1), 1-3.

White, R.T. Indexes used in testing the validity of learning hierarchies. *Journal of Research in Science Teaching*, 1974, 11(1), 61-66.

White, R.T. & Clark, R.C. A test of inclusion which allows errors of measurement. *Psychometrika*, 1973, 38, 77-86.